数理解析研究所講究録
第 1999 巻 2016 年 131-153

CAN properties of the random-coefficient model of demand for nondurable consumer goods in the presence of national micro moments: A simulation study

Yuichiro KANAZAWA, Ph.D.
[1] Faculty of Engineering, Information, and Systems, University of Tsukuba

Kazuhiro NAKAYAMA and Keisuke TAKESHITA
Graduate School of Systems and Information Engineering, University of Tsukuba

Abstract

In this paper, we implement Monte Carlo simulation to examine asymptotic properties of the estimator of random-coefficient logit models of demand for non-durable consumer goods under an equilibrium assumption in the presence of micro moments as the number of the examined regional markets increases. The national micro moments are manufactured from the joint distribution of demographic information of consumers choosing those products with certain discriminating attributes. We observe that adding an equilibrium assumption and the micro moments gives asymptotic normality with sharper asymptotic variance-covariance matrix, while correcting asymptotic bias reported in Freyberger (2015). We discuss possible reasons for such a phenomenon.

[1]This research is supported in part by the Grants-in-Aid for Scientific Research (B) 15H03333, 20310081, and the Grant-in-Aid for Scientific Researchfor Challenging Exploratory Research 25590051 from the Japan Society for the Promotion of Science.

1 Introduction

Industrial organization and some marketing science literature is concerned with the structure of industries in the economy and the behavior of firms and individuals in these industries.

Suppose you observe high prices in an industry. We ask ourselves if this is due to market power, or due to high costs? Unfortunately, however, the important determinants of firm behavior, costs, are usually unobserved. The " new empirical industrial organization" (NEIO; a moniker coined by Bresnahan (1989)) is motivated by this data problem. NEIO takes an indirect approach, in that we obtain estimate of firms' market power expressed by markups by estimating firms' demand functions.

Under NEIO, products are treated as bundles of characteristics and preferences are defined on those characteristics: Each consumer chooses a bundle that maximizes its utility. This characteristics space models as opposed to the more traditional product space models solve " too many parameter " and " new goods " problems assoicated with the latter.

Since consumers have different preferences for different characteristics, and hence make different choices. In this sense, consumers are *heterogeneous*. Thus we need to allow different consumers to have different demographics–income, age, family size, location of residence, and other factors. We then formulate a demand system which is conditional on the consumer ' s characteristics and a vector of parameters which determines the relationship between those characteristics and preferences over products.

To formulate such a demand system using market level data, we proceed as follows. First we draw vectors of consumer characteristics from the distribution of those characteristics, we then determine the choice probability that each of the drawn households would make for a given value of the parameter. Next we aggregate those choice probabilities into a predicted aggregate demand conditional on the parameter vector. Finally we employ a search routine to find the value of that parameter vector that makes these aggregate probabilities as close as possible to the observed market shares. This idea of simulation estimators developed by Pakes (1986) enabled us to "disaggregate"aggregate demand to individual/household heterogeneous behavioral model in principle.

However, there is a problem. Consumer goods are differentiated in many ways. As a result even if we econometricians measured all the relevant characteristics, we could not expect to obtain precise estimates of their impacts. One solution proposed in Berry (1994) is to put in the " important " differentiating characteristics and an unobservable, ξ, which picks up the aggregate effect of the multitude of characteristics that are being omitted. The unobservable ξ was thought to be buried deep inside a highly non-linear set of equations, and hence it was not obvious how to recover it. Berry (1994) showed that there is a unique value for the vector of unobservables that makes the predicted shares exactly equal to the observed shares. Berry, Levinsohn, and Pakes (1995; henceforth BLP(1995)) further provide a contraction mapping technique which transforms the demand system into a system of equations that is linear in these unobservables.

This ξ represents the effect of characteristics that are unknown only to econometricians but not so to consumers as well as producers: Consumers purchases products knowing ξ and producers know ξ when they set prices. As a result, goods that have high values for ξ will be priced higher in any reasonable notion of equilibrium. This produces an analogue to the standard simultaneous equation "endogeneity" or "simultaneity" problem in estimating demand systems in the older demand literature; i.e. prices are correlated with the disturbance term. However, because of the way these unobservables are set up in BLP (1995), we are able to use instruments to formulate generalized method of moments estimation (GMM for short) to overcome this " simultaneity problem. "

Some recent empirical studies in industrial organization and marketing science extend the framework proposed by Berry, Levinsohn and Pakes (1995, henceforth BLP (1995)) by integrating information on consumer demographics into the utility functions in order to make their demand models more realistic and convincing.

Wide availability of public sources of information such as the Current Population Survey (CPS) and the Integrated Public Use Microdata Series (IPUMS) makes these studies possible. Those sources give us information on the joint distribution of the U.S. household's demographics including income, age of household's head, and family size. For example, Nevo (2001)'s examination on price competition in the U.S. ready-to-eat cereal industry uses individual's income, age and a dummy variable indicating if s/he has a child in the utility function. Sudhir (2001) includes household's income to model the U.S. automobile demand in his study of competitive interactions among firms in different market segments.

In analyzing the U.S. automobile market, Petrin (2002) goes further and links demographics of new-vehicle purchasers to characteristics of the vehicles they purchased. Petrin adds a set of functions of the expected value of consumer's demographics given specific product characteristics (e.g. expected family size of households that purchased minivans) as additional moments to the original moments used in BLP (1995) in the GMM estimation. Specifically, he matches the model's probability of new vehicle purchase for different income groups to the observed purchase probabilities in the Consumer Expenditure Survey (CEX) automobile supplement. He also matches model prediction for average household characteristics of vehicle purchasers such as family-size to the data in CEX automobile supplement. Petrin presupposes readily accessible and publicly available market information on the population average.[2] He maintains that "the extra information plays the same role as consumer-level data, allowing estimated substitution patterns and (thus) welfare to directly reflect demographic-driven differences in tastes for observed characteristics" (page, 706, lines 22-25). His intention, it seems, is to reduce the bias associated with "a heavy dependence on the idiosyncratic logit "taste" error" (page 707, lines 5-6).

He explains that "the idea for using these additional moments derives from Imbens and Lancaster (1994). They suggest that aggregate data may contain useful information on the average of micro variables" (page

[2]Berry, Levinsohn, and Pakes (2004), on the other hand, uses detailed consumer-level data, which include not only individuals' choices but also the choices they would have made had their first choice products not been available. Although the proposed method should improve the out-of-sample model's prediction, it requires proprietary consumer-level data, which are not readily available to researchers, as the authors themselves acknowledged in the paper: the CAMIP data "are generally not available to researchers outside of the company" (page 79, line 30).

713, lines 27-29).[3]

It should be noted that these additional moments are subject to simulation and sampling errors in BLP estimation. This is because the expectations of consumer demographics are evaluated conditional on a set of exogenous product characteristics $\boldsymbol{X}$ and an unobserved product quality $\boldsymbol{\xi}$, where the $\boldsymbol{\xi}$ is evaluated with the simulation error induced by BLP's contraction mapping as well as with the sampling error contained in observed market shares. In addition, market information against which the additional moments are evaluated itself contains another type of sampling error. This is because the market information is typically an estimate for the population average demographics obtained from a sample of consumers (e.g. CEX sample), while observed market shares are calculated from another sample of consumers. This error also affects evaluation of the additional moments. In summary each of the four errors (the simulation error, the sampling error in the observed market shares, the sampling error *induced when researcher evaluates the additional moments*, and the sampling error *in the market information itself*) as well as stochastic nature of the product characteristics affects evaluation of the additional moments.

The estimator proposed by Petrin appears to assume that we are able to control impacts from the first four errors. However, it is not apparent if Petrin estimator is consistent and asymptotically normal (CAN) without such a control. Furthermore, it is not known either if how many and in what way individuals need to be sampled in order for Petrin estimator to be more efficient than BLP estimator. Myojo and Kanazawa (2012) formalized Petrin's idea and provide the conditions under which Petrin estimator not only has CAN properties, but is more efficient than BLP estimator as the number of products increases in a national market, a framework designed to analyze durable goods such as automobiles.

On the other hand, packaged goods such as foods, beverages, grooming aids, laundry powder, toiletries and others are sold in a retail outlet whose shelf space is limited. Thus it is unrealistic for manufacturer to increase the number of products and expect retailers to carry all of them. In addition, shoppers for those packaged goods are likely to shop locally, so their prices are likely to be determined in the local market. Each one of these markets, of course, reflects regional or local demographics, and this may induce certain product characteristic(s) to be valued in one local/regional market, but not so much in the other. As a result, consumers in different markets may perceive the same product somewhat differently. To capture their heterogeneity for those product categories, we need to be able to have a tool to analyze circumstances where the number of products is fixed, but we are able to observe many of these local/regional markets.

Freyberger (2015) develops asymptotic theory for estimated parameters in differentiated product "demand" systems with a small number of products and a large number of markets T. His asymptotic theory takes into account the fact that "the predicted market shares are approximated by Monte Carlo integration with R draws and that the observed market shares are approximated from a sample of N consumers." As expected, he found "both approximations affect the asymptotic distribution, because they both lead to a

[3]Original intension of Imbens and Lancaster is to improve efficiency for a class of the extremum estimators. There is a difference between Petrin's and Imbens and Lancaster's approaches in sampling process to construct original and additional sample moments. Petrin combines the sample moments calculated *over products* with additional moments calculated *over individuals*, while Imbens and Lancaster use the moments calculated over the same individuals.

bias and a variance term in the asymptotic expansion of the estimator." He showed that "when R and N do not increase faster than the number of markets, the bias terms dominate the variance terms, and the asymptotic distribution might not be centered at 0 and standard confidence intervals do not have the right size, even asymptotically." He showed that the leading bias terms can be eliminated by using an analytic bias correction method.

In this paper, we also deal with the same circumstances as Freyberger (2015) in which the number of products is fixed while the number of markets increases. We introduce *the pricing equation* as well as *the additional micro moments or summary statistics that provide information on the joint distribution of consumer demographics and product characteristics* as Petrin (2002) and Myojo and Kanazawa (2012) did to the model and investigate if the accuracy, both in terms of bias and variance, of the estimator is improved upon via Monte Carlo simulation. This study is designed to facilitate theoretical examination of the problem.

2 System of Demand and Supply with Additional Moments in Multiple Markets

In this section, we define the product space precisely, and reframe the estimation procedure of BLP (1995) when combining the demand and supply side moments with the additional moments relating consumer demographics to the characteristics of products they purchase. For convenience in comparison to Berry, Linton, and Pakes (2004) (BLP(2004) for short), notation and most definitions are kept as same as possible to those in BLP (2004).

2.1 Demand Side Model

Regional markets are indexed by $m = 1, \ldots, M$ and the population in market M is I_m. We assume the same finite set of products, $j = 1, \ldots, J$, is available in each regional market. We also assume that each consumer only participates in one market and chooses one product including "outside" good that maximizes his/her utility within that market. We assume the utility of consumer i for product j in regional market m to be

$$\begin{aligned} u_{ij}^m &= \alpha \ln(y_i^m - p_j^m) + x_j'\beta + \xi_j^m + \textstyle\sum_{k=1}^K \pi_k x_{jk} \nu_{ik}^m + \epsilon_{ij}^m \\ &= \alpha \ln(y_i^m - p_j^m) + \textstyle\sum_{k=1}^K x_{jk}(\beta_k + \pi_k \nu_{ik}^m) + \xi_j^m + \epsilon_{ij}^m. \end{aligned} \tag{1}$$

This specification closely parallels with that of BLP (1995) except that some demographic and product characteristics are indexed by m as well. For instance, y_i^m is the income of consumer i, p_j^m is the price of product j, ξ_j^m is the unobserved product characteristics for product j, ν_{ik}^m is consumer i's taste for k-th product characteristic, and ϵ_{ij}^m is unobserved idiosyncratic tastes of consumer i, all in market m. We assume that ϵ_{ij}^m are i.i.d with type I extreme value.

The price and the unobserved product characteristic of the same product may as well vary from one market to the other because of the differences in demographics between markets. However, we assume the parameters $(\alpha, \boldsymbol{\beta})$ are not indexed by m.

Although most observed product characteristics are not correlated with the unobserved product characteristics $\xi_j^m \in \Re$, $j = 1, \ldots, J$, some of them (e.g. price) are. We denote the vector of observed product characteristics $\boldsymbol{x}_j = (\boldsymbol{x}_{1j}', \boldsymbol{x}_{2j}')'$ where $\boldsymbol{x}_{1j} \in \Re^{K_1}$ are exogenous and not correlated with ξ_j^m, while $\boldsymbol{x}_{2j} \in \Re^{K_2}$ are endogenous and correlated with ξ_j^m where $K = K_1 + K_2$. Observed product characteristics other than price, denoted as $\boldsymbol{x}_{1j}$ are not indexed by m because we assume that the same set of products is sold in every market. We assume the set of exogenous product characteristics $(\boldsymbol{x}_{1j}, \xi_j^1, \xi_j^2, \ldots, \xi_j^M)$, $j = 1, \ldots, J$ is a random sample from the underlying population of product characteristics, and is thus independent across j similar to the framework of BLP (1995).

The ξ_j^m's are assumed to be mean independent of $\boldsymbol{X}_1 = (\boldsymbol{x}_{11}, \ldots, \boldsymbol{x}_{1J})'$

$$\mathrm{E}_{\xi^m|X_1}[\xi_j^m|\boldsymbol{X}_1] = 0 \tag{2}$$

with probability 1. We also assume the conditional variance of ξ_j^m on $\boldsymbol{x}_{1j}$ is finite

$$\sup_{1\le m\le M} \max_{1\le j\le J} \mathrm{E}_{\xi^m|x_{1j}}[(\xi_j^m)^2|\boldsymbol{x}_{1j}] < \infty$$

with probability one. We denote by $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_J)'$ the set of all observed product characteristics.

We define the utility for outside good as

$$u_{i0}^m = \alpha \ln(y_i^m) + \epsilon_{i0}^m,$$

and redefine the utility as difference from outside good, $U_{ij}^m = u_{ij}^m - u_{i0}^m$ as

$$\begin{aligned} U_{ij}^m &= u_{ij}^m - u_{i0}^m \\ &= \alpha \ln(y_i^m - p_j^m) + \sum_{k=1}^{K} x_{jk}(\beta_k + \pi_k \nu_{ik}^m) + \xi_j^m + \epsilon_{ij}^m - (\alpha \ln(y_i^m) + \epsilon_{i0}^m) \\ &= \alpha \ln\left(1 - \frac{p_j^m}{y_i^m}\right) + \boldsymbol{x}_j'\boldsymbol{\beta} + \sum_{k=1}^{K} x_{jk}\pi_k \nu_{ik}^m + \xi_j^m + \epsilon_{ij}^m - \epsilon_{i0}^m \\ &= \delta_j^m(\boldsymbol{x}_j, \xi_j^m; \boldsymbol{\beta}) + \mu_{ij}^m(\boldsymbol{x}_j, p_j^m, y_i^m, \boldsymbol{\nu}_i^m; \alpha, \boldsymbol{\pi}) + \epsilon_{ij}^m - \epsilon_{i0}^m, \end{aligned} \tag{3}$$

where

$$\begin{aligned} \delta_j^m(\boldsymbol{x}_j, \xi_j^m; \boldsymbol{\beta}) &= \boldsymbol{x}_j'\boldsymbol{\beta} + \xi_j^m, \\ \mu_{ij}^m(\boldsymbol{x}_j, p_j^m, y_i^m, \boldsymbol{\nu}_i^m; \alpha, \boldsymbol{\pi}) &= \alpha \ln\left(1 - \frac{p_j^m}{y_i^m}\right) + \sum_{k=1}^{K} x_{jk}\pi_k \nu_{ik}^m. \end{aligned}$$

This redefinition standardizes the utility function so that $U_{i0} = 0$. Note that $\delta_j^m(\boldsymbol{x}_j, \xi_j^m; \boldsymbol{\beta})$ depends on product characteristics, but not on consumer demographics, while $\mu_{ij}^m(\boldsymbol{x}_j, p_j^m, y_i^m, \boldsymbol{\nu}_i^m; \alpha, \boldsymbol{\pi})$ depends on both.

The conditional probability σ_{ij}^m for consumer i in market m to choose product j

$$\sigma_{ij}^m(\boldsymbol{X}, \boldsymbol{\xi}^m, y_i^m, \boldsymbol{\nu}_i^m; \boldsymbol{\theta}_d) = \frac{\exp(\delta_j^m + \mu_{ij}^m)}{\sum_{j=0}^J \exp(\delta_j^m + \mu_{ij}^m)}, \tag{4}$$

is a map from the set of observed and unobserved product characteristics $\boldsymbol{X}$ and $\boldsymbol{\xi}^m = (\xi_1^m, \ldots, \xi_J^m)'$, demographics y_i^m and tastes $\nu_i^m \in \Re^\nu$ of consumer i in market m, and a demand parameter vector $\boldsymbol{\theta}_d \in \Theta_d$. We assume $\sigma_{ij}^m(\boldsymbol{X}, \boldsymbol{\xi}^m, y_i^m, \boldsymbol{\nu}_i^m; \boldsymbol{\theta}_d) > 0$ for all possible values of $(\boldsymbol{X}, \boldsymbol{\xi}^m, y_i^m, \boldsymbol{\nu}_i^m; \boldsymbol{\theta}_d)$.

The BLP framework generates the vector $\boldsymbol{\sigma}^m$ of market shares in region m by aggregating over the individual choice probability σ_{ij}^m over the joint distribution $P^m(\cdot)$ of the consumer demographics and tastes $(y_i^m, \boldsymbol{\nu}_i^m)$ as

$$\sigma_j^m(\boldsymbol{X}, \boldsymbol{\xi}^m, \boldsymbol{\theta}_d, P^m) = \int \sigma_{ij}^m(\boldsymbol{X}, \boldsymbol{\xi}^m, y_i^m, \boldsymbol{\nu}_i^m; \boldsymbol{\theta}_d) dP^m(y_i^m, \boldsymbol{\nu}_i^m) \tag{5}$$

where P^m is typically the empirical joint distribution of the consumer demographics and tastes from a random sample drawn from the underlying population joint distribution $P^{m,0}$ of demographics and tastes in market m. If we evaluate (5) at $(\boldsymbol{\theta}_d^0, P^{m,0})$, where $\boldsymbol{\theta}_d^0$ is the true value of demand side parameters, we have the "conditionally true" market shares $\boldsymbol{s}^{m,0}$ given the product characteristics $(\boldsymbol{X}, \boldsymbol{\xi}^m)$ in market m. That is,

$$\boldsymbol{\sigma}^m(\boldsymbol{X}, \boldsymbol{\xi}^m, \boldsymbol{\theta}_d^0, P^{m,0}) \equiv \boldsymbol{s}^{m,0}. \tag{6}$$

where $\boldsymbol{s}^{m,0} = (s_1^{m,0}, \ldots, s_J^{m,0})'$.

If we solve the identity (6) at any $(\boldsymbol{\theta}_d, \boldsymbol{X}, \boldsymbol{s}^m, P^m) \neq (\boldsymbol{\theta}_d^0, \boldsymbol{X}, \boldsymbol{s}^{m,0}, P^{m,0})$, the independence assumption for the resulting $\xi_j^m(\boldsymbol{\theta}_d, \boldsymbol{X}, s_j^m, P^m)$ no longer holds because the two factors deciding the ξ_j^m—the market share s_j^m and the endogenous product characteristics $\boldsymbol{x}_{2j}^m$ for product j—are endogenously determined through the market equilibrium as a function of the characteristics of all products. However, if we solve (6) at $(\boldsymbol{\theta}_d, \boldsymbol{X}, \boldsymbol{s}^m, P^m) = (\boldsymbol{\theta}_d^0, \boldsymbol{X}, \boldsymbol{s}^{m,0}, P^{m,0})$, we are able to retrieve the original $\boldsymbol{\xi}^m = \boldsymbol{\xi}^m(\boldsymbol{\theta}_d^0, \boldsymbol{X}, \boldsymbol{s}^{m,0}, P^{m,0})$ for $m = 1, \ldots, M$ and they are independent. [4]

2.2 Supply Side Model

The supply side model formulates the pricing equations for the J products marketed. We assume an oligopolistic market where a finite number F of "national" suppliers ($f = 1, \ldots, F$), and each supplier produces $\mathcal{J}_f$ products. Suppliers are maximizers of profit from the combination of products they produce. Assuming Bertrand-Nash pricing provides the first order condition for the product $j \in \mathcal{J}_f$ of the manufac-

[4] BLP (1995) provides general conditions under which there is a unique solution $\boldsymbol{\xi}$ for

$$\boldsymbol{s} - \boldsymbol{\sigma}(\boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P) = \boldsymbol{0} \tag{7}$$

for every $(\boldsymbol{X}, \boldsymbol{\theta}_d, \boldsymbol{s}, P) \in \mathcal{X} \times \Theta_d \times \mathcal{S}_J \times \mathcal{P}$, where $\mathcal{X}$ is a space for the product characteristics $\boldsymbol{X}$, $\mathcal{S}_J$ is a space for the market share vector $\boldsymbol{s}$, and $\mathcal{P}$ is a family of probability measures.

turer f in regional market m is

$$\sigma_j^m(\boldsymbol{X}, \boldsymbol{p}^m, \boldsymbol{\xi}^m, \boldsymbol{\theta}_d, P^m) + \sum_{h \in \mathcal{J}_f} (p_h^m - mc_h^m) \frac{\partial \sigma_h^m(\boldsymbol{X}, \boldsymbol{p}^m, \boldsymbol{\xi}^m, \boldsymbol{\theta}_d, P^m)}{\partial p_j^m} = 0, \tag{8}$$

where $\boldsymbol{p}^m = (p_1^m, \ldots, p_J^m)$. This equation can be expressed in matrix form as

$$\boldsymbol{\sigma}^m(\boldsymbol{X}, \boldsymbol{\xi}^m, \boldsymbol{\theta}_d, P^m) + \boldsymbol{\Delta}^m(\boldsymbol{p}^m - \boldsymbol{mc}^m) = \boldsymbol{0} \tag{9}$$

where $\boldsymbol{\Delta}^m$ is the $J \times J$ non-singular gradient matrix whose (j, h) element is defined by

$$\Delta_{jh}^m = \begin{cases} \partial \sigma_h^m(\boldsymbol{X}, \boldsymbol{\xi}^m, \boldsymbol{\theta}_d, P^m)/\partial p_j^m, & \text{if the products } j \text{ and } h \text{ are produced by the same firm,} \\ 0, & \text{otherwise.} \end{cases}$$

We define the marginal cost mc_j^m as an implicit function of the observed cost shifters $\boldsymbol{w}_j$ common to all the regional markets and the unobserved cost shifter ω_j^m that depends on the market as

$$g(mc_j^m) = \boldsymbol{w}_j' \boldsymbol{\theta}_c + \omega_j^m, \tag{10}$$

where g is a monotonic function and $\boldsymbol{\theta}_c \in \Theta_c$ is a vector of cost parameters.

While the choice of $g(\cdot)$ depends on the application, we assume $g(\cdot)$ is continuously differentiable with a finite derivative for all realizable values of cost. Suppose that the observed cost shifters $\boldsymbol{w}_j$ consist of exogenous $\boldsymbol{w}_{1j} \in \Re^{L_1}$ as well as endogenous $\boldsymbol{w}_{2j} \in \Re^{L_2}$, and thus we write $\boldsymbol{w}_j = (\boldsymbol{w}_{1j}', \boldsymbol{w}_{2j}')'$ and $\boldsymbol{W} = (\boldsymbol{w}_1, \ldots, \boldsymbol{w}_J)'$. The exogenous cost shifters include not only the cost variables determined outside the market under consideration (e.g. factor price), but also the product design characteristics suppliers cannot immediately change in response to fluctuation in demand. The cost variables determined by the market equilibrium (e.g. production scale) are treated as endogenous cost shifters.

As in the formulation of $(\boldsymbol{x}_{1j}, \xi_j^1, \xi_j^2, \ldots, \xi_j^M)$, $j = 1, \ldots, J$ on the demand side, we assume the set of exogenous cost shifters $(\boldsymbol{w}_{1j}, \omega_j^1, \omega_j^2, \ldots, \omega_j^M)$, $j = 1, \ldots, J$ is a random sample from the underlying population of cost shifters. Thus $(\boldsymbol{w}_{1j}, \omega_j^1, \omega_j^2, \ldots, \omega_j^M)$ are assumed to be independent across j, while $\boldsymbol{w}_{2j}$ are in general not independent across j as they are determined in the market as functions of cost shifters of other products. Similar to the demand side unobservables, the unobserved cost shifter ω_j is assumed to be mean independent of the exogenous cost shifters $\mathbf{W}_1 = (\boldsymbol{w}_{11}, \ldots, \boldsymbol{w}_{1J})'$, and satisfy with probability one,

$$\mathrm{E}_{\omega^m|X_1}[\omega_j^m|\boldsymbol{W}_1] = 0, \quad \text{and} \quad \sup_{1 \le m \le M} \max_{1 \le j \le J} \mathrm{E}_{\omega^m|X_1}[(\omega_j^m)^2|\boldsymbol{w}_{1j}] < \infty. \tag{11}$$

Define $\boldsymbol{g}(\boldsymbol{x}) \equiv (g(x_1), \ldots, g(x_J))$. Solving the first order condition (9) with respect to $\boldsymbol{mc}^m$ and substituting for (10) give the vector of the unobserved cost shifters

$$\boldsymbol{\omega}^m(\boldsymbol{\theta}, \mathrm{s}^m, P^m) = \boldsymbol{g}(\boldsymbol{p}^m + (\boldsymbol{\Delta}^m)^{-1}\boldsymbol{\sigma}^m) - \boldsymbol{W}\boldsymbol{\theta}_c. \tag{12}$$

Notice that the parameter vector $\boldsymbol{\theta}$ in ω contains both the demand and supply parameters, i.e. $\boldsymbol{\theta} = (\boldsymbol{\theta}_d', \boldsymbol{\theta}_c')'$. Since the profit margin $m_{g_j}(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P)$ for product j is determined not only by its unobserved product characteristics ξ_j, but by those of the other products in the market, these ω_j are in general dependent across j when $(\boldsymbol{\theta}, \boldsymbol{s}, P) \neq (\boldsymbol{\theta}^0, \boldsymbol{s}^0, P^0)$. However, when (12) is evaluated at $(\boldsymbol{\theta}, \boldsymbol{s}, P) = (\boldsymbol{\theta}^0, \boldsymbol{s}^0, P^0)$, we are able to recover the original $\omega_j, j = 1, \ldots, J$, and they are assumed independent across j and m.

2.3 GMM Estimation with National Micro Moments

Let us define the $J \times L_d$ demand side instrument matrix $\boldsymbol{Z}_d = (\boldsymbol{z}_1^d, \ldots, \boldsymbol{z}_J^d)'$ whose components $\boldsymbol{z}_j^d$ can be written as $\boldsymbol{z}_j^d(\boldsymbol{x}_{11}, \ldots, \boldsymbol{x}_{1J}) \in \Re^{L_d}$, where $\boldsymbol{z}_j^d(\cdot) : \Re^{K_1 \times J} \to \Re^{L_d}$ for $j = 1, \ldots, J$. It should be noted that the demand side instruments $\boldsymbol{z}_j^d$ for product j are assumed to be a function of the exogenous characteristics not only of its own, but of the other products in the market. This is because the instruments by definition must correlate with the product characteristics $\boldsymbol{x}_{2j}$, and these endogenous variables $\boldsymbol{x}_{2j}$ (e.g. price) are determined by both its own and its competitors' product characteristics.

Similar to the demand side, we define the $J \times L_c$ supply side instruments $\boldsymbol{Z}_c = (\boldsymbol{z}_1^c, \ldots, \boldsymbol{z}_J^c)'$ as a function of the exogenous cost shifters $(\boldsymbol{w}_{11}, \ldots, \boldsymbol{w}_{1J})$ of all the products. Here, $\boldsymbol{z}_j^c(\boldsymbol{w}_{11}, \ldots, \boldsymbol{w}_{1J}) \in \Re^{L_c}$ and $\boldsymbol{z}_j^c(\cdot) : \Re^{L_1 \times J} \to \Re^{L_c}$ for $j = 1, \ldots, J$. We also note that some of the exogenous product characteristics $\boldsymbol{x}_{1j}$ affect the price of the product because they affect manufacturing cost. Thus those $\boldsymbol{x}_{1j}$ may be included among the exogenous cost shifters $\boldsymbol{w}_{1j}$ if they are uncorrelated with the unobservable cost shifter ω_j.

Assume, for moment, that we know the underlying taste distribution of P^0 and that we are able to observe the true market share $\boldsymbol{s}^0$. Considering stochastic nature of the product characteristics $\boldsymbol{X}_1$ and $\boldsymbol{\xi}$, we set forth the demand side restriction as

(13)
$$\mathrm{E}_{\mathrm{x}_1, \xi^m}\left[\boldsymbol{z}_j^d \xi_j^m(\boldsymbol{\theta}_d, \boldsymbol{X}, \boldsymbol{s}^{m,0}, P^{m,0})\right] = \boldsymbol{0}$$

at $\boldsymbol{\theta}_d = \boldsymbol{\theta}_d^0$ where the expectation is taken with respect not only to $\boldsymbol{\xi}$, but also to $\boldsymbol{X}_1$ for $m = 1, \ldots, M$. The supply side restriction we use is

(14)
$$\mathrm{E}_{\mathrm{w}_1, \omega^m}\left[\boldsymbol{z}_j^c \omega_j^m(\boldsymbol{\theta}, \boldsymbol{X}, \boldsymbol{s}^{m,0}, P^{m,0})\right] = \boldsymbol{0}$$

at $\boldsymbol{\theta} = \boldsymbol{\theta}^0$ for $m = 1, \ldots, M$. We extend the BLP framework to use the orthogonality conditions between the unobserved product characteristics (ξ_j^m, ω_j^m) and the exogenous instrumental variables $(\boldsymbol{z}_j^d, \boldsymbol{z}_j^c)$ as moment conditions to obtain the GMM estimate of the parameter $\boldsymbol{\theta}$. The "sample" moments for the demand and supply systems are

(15)
$$\boldsymbol{G}_M(\boldsymbol{\theta}, \boldsymbol{X}, \boldsymbol{s}^0, P^0) = \begin{pmatrix} \boldsymbol{G}_M^d(\boldsymbol{\theta}_d, \boldsymbol{X}, \boldsymbol{s}^0, P^0) \\ \boldsymbol{G}_M^c(\boldsymbol{\theta}, \boldsymbol{X}, \boldsymbol{s}^0, P^0) \end{pmatrix} = \begin{pmatrix} \sum_{m=1}^M \sum_{j=1}^J \boldsymbol{z}_j^d \xi_j^m(\boldsymbol{\theta}_d, \boldsymbol{X}, \boldsymbol{s}^0, P^0)/M \\ \sum_{m=1}^M \sum_{j=1}^J \boldsymbol{z}_j^c \omega_j^m(\boldsymbol{\theta}, \boldsymbol{X}, \boldsymbol{s}^0, P^0)/M \end{pmatrix}.$$

It should be noted that we are interested in circumstances where the number J of products in region m is fixed, while the number M of markets increases and this requires averaging of $\sum_{j=1}^{J} z_j^m \xi_j^m$ over m thanks to the assumption that ξ_j^m, $j = 1, \ldots, J$, $m = 1, \ldots, M$ are i.i.d when conditional on $\mathbf{X}_1$.

For some markets, market summaries such as average demographics of consumers who purchased a specific type of product are publicly available, even if detailed individual-level data such as purchasing histories are not. We now operationalize the idea put forth by Petrin (2002), which extends the BLP framework by adding moment conditions constructed from the market summary data. First we require a few definitions. A *discriminating attribute* is an observable product characteristic of products that determines a subset of products in the market, those products that possess the attribute. We denote the set of products with discriminating attribute q as $\mathcal{Q}_q$, and the consumer's choice as C_i. We will write "consumer i chooses discriminating attribute q" when $C_i \in \mathcal{Q}_q$. We assume there is a finite number of discriminating attributes $q = 1, \ldots, N_p$, and that the market share of each discriminating attribute is positive, i.e., $\Pr[C_i \in \mathcal{Q}_q | \boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0)] > 0$ for all q in $1, \ldots, N_p$.

We next consider the expectation of a consumer's demographics conditional on a specific discriminating attribute. Suppose that consumer i's demographics can be decomposed into observable and unobservable components $\boldsymbol{\nu}_i = (\boldsymbol{\nu}_i^{obs}, \boldsymbol{\nu}_i^{unobs})$. The densities of $\boldsymbol{\nu}_i$ and $\boldsymbol{\nu}_i^{obs}$ are respectively denoted as $P^0(d\boldsymbol{\nu}_i)$ and $P^0(d\boldsymbol{\nu}_i^{obs})$. Some observable demographic variables such as age, family size, or income, are already numerical, but other demographics such as household with children, belonging to a certain age group, choice of residential area, must be numerically expressed using indicators. We denote this numerically represented D dimensional demographics as $\boldsymbol{\nu}_i^{obs} = (\nu_{i1}^{obs}, \ldots, \nu_{iD}^{obs})'$. We assume that the joint density of demographics $\boldsymbol{\nu}_i^{obs}$ is of bounded support. Consumer i's d-th observed demographic $\nu_{id}^{obs}, d = 1, \ldots, D$ is averaged over all consumers choosing discriminating attribute q in the population to obtain the conditional expectation $\eta_{dq}^0 = \mathrm{E}[\nu_{id}^{obs} | C_i \in \mathcal{Q}_q, \boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0)]$.

We assume η_{dq}^0 has a finite mean and variance for all J, i.e. $\mathrm{E}_{\mathrm{x},\xi}[\eta_{dq}^0] < \infty$ and $\mathrm{V}_{\mathrm{x},\xi}[\eta_{dq}^0] < \infty$ for $d = 1, \ldots, D, q = 1, \ldots, N_p$.

Let $\Pr[d\nu_{id}^{obs} | C_i \in \mathcal{Q}_q, \boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0)]$ be the conditional density of consumer i's demographics ν_{id}^{obs} given his/her choice of discriminating attribute q and product characteristics $(\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0))$. Since the conditional expectation η_{dq}^0 can be written as

$$
\begin{aligned}
(16) \qquad & \mathrm{E}[\nu_{id}^{obs} | C_i \in \mathcal{Q}_q, \boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0)] \\
&= \int \nu_{id}^{obs} \Pr[d\nu_{id}^{obs} | C_i \in \mathcal{Q}_q, \boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0)] \\
&= \frac{\int \nu_{id}^{obs} \Pr[C_i \in \mathcal{Q}_q | \boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0), \nu_{id}^{obs}] P^0(d\nu_{id}^{obs})}{\Pr[C_i \in \mathcal{Q}_q | \boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0)]} \\
&= \frac{\int \nu_{id}^{obs} \Pr[C_i \in \mathcal{Q}_q | \boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0), \boldsymbol{\nu}_i] P^0(d\boldsymbol{\nu}_i)}{\Pr[C_i \in \mathcal{Q}_q | \boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0)]} \\
&= \int \nu_{id}^{obs} \frac{\sum_{j \in \mathcal{Q}_q} \sigma_{ij}(\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0), \boldsymbol{\nu}_i; \boldsymbol{\theta}_d)}{\sum_{j \in \mathcal{Q}_q} \sigma_j(\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d, P^0)} P^0(d\boldsymbol{\nu}_i),
\end{aligned}
$$

we can form an identity, which is the basis for additional moment conditions

$$\eta^0_{dq} - \int \nu^{obs}_{id} \frac{\sum_{j\in Q_q} \sigma_{ij}(\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}^0_d, \boldsymbol{s}^0, P^0), \boldsymbol{\nu}_i; \boldsymbol{\theta}^0_d)}{\sum_{j\in Q_q} \sigma_j(\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}^0_d, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}^0_d, P^0)} P^0(d\boldsymbol{\nu}_i) \equiv 0 \tag{17}$$

for $q = 1, \dots, N_p, d = 1, \dots, D$.

Although P^0 is so far assumed known, we typically are not able to calculate the second term on the left hand side of (17) analytically and will have to approximate it by using the empirical distribution P^T of an i.i.d. sample $\boldsymbol{\nu}_t, t = 1, \dots, T$ from the underlying distribution P^0. The corresponding sample moments $\boldsymbol{G}^a_{J,T}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0, \boldsymbol{\eta}^0)$, where superscript a stands for "additional," are

$$\boldsymbol{G}^a_{M,T}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0, \boldsymbol{\eta}^0) = \boldsymbol{\eta}^0 - \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{\nu}^{obs}_t \otimes \boldsymbol{\psi}_t(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d, P^0) \tag{18}$$

where

$$\boldsymbol{\eta}^0 = (\eta^0_{11}, \dots, \eta^0_{1N_p}, \dots, \eta^0_{D1}, \dots, \eta^0_{DN_p})', \qquad \boldsymbol{\psi}_t(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P) = \begin{pmatrix} \frac{\sum_{j\in Q_1} \sigma_{tj}(\boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{\nu}_t; \boldsymbol{\theta}_d)}{\sum_{j\in Q_1} \sigma_j(\boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P)} \\ \vdots \\ \frac{\sum_{j\in Q_{N_p}} \sigma_{tj}(\boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{\nu}_t; \boldsymbol{\theta}_d)}{\sum_{j\in Q_{N_p}} \sigma_j(\boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P)} \end{pmatrix}. \tag{19}$$

The symbol $\otimes$ denotes the Kronecker product. The quantity $\boldsymbol{\psi}_t(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P)$ is consumer t's model-calculated probability of purchasing products with discriminating attribute q relative to the model-calculated market share of the same products. Note that these additional moments are again conditional on product characteristics $(\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0))$, and thus depend on the sample sizes J and T.

We use the set of the three moments, two from (16) and from (18) as

$$\boldsymbol{G}_{M,T}(\boldsymbol{\theta}, \boldsymbol{X}, \boldsymbol{s}^0, P^0, \boldsymbol{\eta}^0) = \begin{pmatrix} \boldsymbol{G}^d_M(\boldsymbol{\theta}_d, \boldsymbol{X}, \boldsymbol{s}^0, P^0) \\ \boldsymbol{G}^c_M(\boldsymbol{\theta}, \boldsymbol{X}, \boldsymbol{s}^0, P^0) \\ \boldsymbol{G}^a_{M,T}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0, \boldsymbol{\eta}^0) \end{pmatrix} \tag{20}$$

to estimate $\boldsymbol{\theta}$.

As pointed out in BLP (2004), we have two issues when evaluating $||\boldsymbol{G}_{J,T}(\boldsymbol{\theta}, \boldsymbol{s}^0, P^0, \boldsymbol{\eta}^0)||$. First, although we assume $P^{m,0}$ is known, we typically are not able to calculate $\boldsymbol{\sigma}(\boldsymbol{X}, \boldsymbol{\xi}^m, \boldsymbol{\theta}_d, P^{m,0})$ analytically and have to approximate it by a simulator, say $\boldsymbol{\sigma}(\boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P^{m,R_m})$, where P^{m,R_m} is the empirical measure of an i.i.d. sample $(y^m_r, \boldsymbol{\nu}^m_r)$, $r = 1, \dots, R_m$ in market m from the underlying true distribution $P^{m,0}$ in market m. The sample $\boldsymbol{\nu}_r, r = 1, \dots, R$ is assumed independent of the sample $\boldsymbol{\nu}_t, t = 1, \dots, T$ in (18) for evaluating the additional moments. The simulated market shares are then given by

(21)

$$\sigma_j(\boldsymbol{X}, \boldsymbol{\xi}^m, \boldsymbol{\theta}_d, P^{m,R_m}) = \int \sigma_{ij}(\boldsymbol{X}, \boldsymbol{\xi}^m, y^m_i, \boldsymbol{\nu}^m_i; \boldsymbol{\theta}_d) dP^{m,R_m}(y^m_i, \boldsymbol{\nu}_i) \equiv \frac{1}{R_m}\sum_{r=1}^{R_m} \sigma_{rj}(\boldsymbol{X}, \boldsymbol{\xi}^m, y^m_r, \boldsymbol{\nu}^m_r; \boldsymbol{\theta}_d).$$

Second, we are not necessarily able to observe the true market shares $s^{m,0}$ in market m. Instead, the vector of observed market shares, s^{m,n_m}, are typically constructed from n_m i.i.d. draws from the population of consumers, and hence is not equal to the population value $s^{m,0}$ in general. The observed market share of product j in market m is

$$s_j^{m,n_m} = \frac{1}{n_m}\sum_{i=1}^{n_m} 1(C_i = j), \tag{22}$$

where the indicator variable $1(C_i = j)$ takes value 1 if $C_i = j$ and 0 otherwise. Since C_i denotes the choice of randomly sampled consumer i, they are i.i.d. across i.

We substitute $\xi_j^m(\boldsymbol{\theta}_d, \boldsymbol{X}, s^{m,n_m}, P^{m,R_m})$, $m = 1, \ldots, M$, the solution of $s^{m,n_m} = \boldsymbol{\sigma}(\boldsymbol{X}, \boldsymbol{\xi}^m, \boldsymbol{\theta}_d, P^{m,R_m})$ for $\xi_j(\boldsymbol{\theta}_d, \boldsymbol{X}, s^0, P^0)$ in (16) to obtain

$$\boldsymbol{G}_M^d\left(\boldsymbol{\theta}_d, \{s^m\}, \{P^{R_m}\}\right) = \sum_{m=1}^{M} w_m \left\{\sum_{j=1}^{J} \boldsymbol{z}_j^d \xi_j^m(\boldsymbol{\theta}_d, \boldsymbol{X}, s^{m,n_m}, P^{m,R_m})\right\}, \tag{23}$$

where w_m is non-stochastic weight summed up to 1, such as market size.

Similarly, for the supply side, we construct the sample analogue of the regional orthogonality conditions for supply side as follows.

$$\boldsymbol{G}_M^c\left(\boldsymbol{\theta}, \{s^m\}, \{P^{R_m}\}\right) = \sum_{m=1}^{M} w_m \left\{\frac{1}{J}\sum_{j=1}^{J} \boldsymbol{z}_j^c \omega_j^m(\boldsymbol{\theta}, \boldsymbol{X}, \boldsymbol{W}, s^{m,n_m}, P^{m,R_m})\right\}. \tag{24}$$

Suppose that a national micro moments are avialble as in Petrin (2002) and we assume the econometrician obtains the national population database P, not regional databases $P^m (m = 1, \ldots, M)$. In general, we do not know the conditional expectation of demographics η_{dq}^0. Instead, we have an estimate η_{dq}^N from independent source, typically generated from a sample of N consumers. The sample counterparts to (18) for the additional moments are thus

$$\boldsymbol{G}_{M,T}^a(\boldsymbol{\theta}_d, \{s^m\}, \{\boldsymbol{P}^{R_m}\}, \boldsymbol{\eta}^N) = \boldsymbol{\eta}^N - \frac{1}{T}\sum_{t=1}^{T} \nu_t^{m,obs} \otimes \boldsymbol{\psi}^m(\boldsymbol{\xi}^m(\boldsymbol{\theta}_d, s^{n_m}, \boldsymbol{P}^{R_m}), \boldsymbol{\theta}_d, \boldsymbol{P}^{R_m}) \tag{25}$$

where the symbol $\otimes$ denotes the Kronecker product and

$$\begin{aligned} \boldsymbol{\eta}^N &= (\eta_{11}^N, \ldots, \eta_{1N_p}^N, \ldots, \eta_{D1}^N, \ldots, \eta_{DN_p}^N)', \\ \boldsymbol{\psi}_t^m(\boldsymbol{\xi}, \boldsymbol{\theta}, \boldsymbol{P}) &= \begin{pmatrix} \frac{\sum_{j \in Q_1} \sigma_{tj}^m(\boldsymbol{X}, \boldsymbol{\xi}, \nu_t^m; \boldsymbol{\theta}_d)}{\sum_{j \in Q_1} \sigma_j(\boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, \boldsymbol{P})} \\ \vdots \\ \frac{\sum_{j \in Q_{N_p}} \sigma_{tj}^m(\boldsymbol{X}, \boldsymbol{\xi}, \nu_t^m; \boldsymbol{\theta}_d)}{\sum_{j \in Q_{N_p}} \sigma_j(\boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, \boldsymbol{P})} \end{pmatrix}. \end{aligned}$$

3 Monte Carlo Experiments

We now examine the effect of adding pricing equation as well as additional micro moments in the multiple market demand and supply system by Monte Carlo experiments. Since we consider the case where one national market is divided into multiple M markets, as the number of markets grows larger, we obtain the more detailed information on the products and consumers. We want to know whether the additional moment conditions only on "national" consumer information works in this situation. Further, we check the needed order of sample size T relative to the number of markets to show CAN properties. We do not check the effect of sample size R_m in this paper. Note that sample size R_m, $m = 1, \dots, M$, and T are chosen by the econometrician. [5].

3.1 Primary Settings of the Simulation

In this subsection, the primary settings of the simulation are shown. In the simulation study, the utility function of consumer i for product j in regional market m is specified as

$$u_{ij}^m = -\alpha p_j^m + \beta_0 x_j + \beta_1 x_j \nu_i^m + \xi_j^m + \epsilon_{ij}^m. \tag{26}$$

The observed product characteristics x_j and unobserved product characteristics ξ_j^m are random draws from $N(3,1)$ and $N(0,1)$ respectively and x_j and ξ_m^j are independently drawn. The consumer demographics ν_i^m is random draws from $N(0,1)$ and consumer's idiosyncratic tastes term ϵ_{ij}^m is assumed to be i.i.d. with extreme value to derive logit model.

The prices p_j^m of product j in market m, $j = 1, \dots, J$, $m = 1, \dots, M$, is endogenously determined in each regional market equilibrium, so differ from market to market by solving (9) with Newton-Raphson method. The price is the only endogenous variable in this experiment. We set the true demand side parameters as $\alpha = 1.0$, $\beta_0 = 1.0$ and $\beta_1 = 0.5$ common to all regional markets.

We set the total number of consumers in the national market to be $I = 10,000$ and there exist the same number I/M of consumers in each regional market. We need the population of size I because we must construct the true market share. The weight for each regional market m is $w_m = l_m = 1/M$ common to all the regional markets. We draw R_m or T consumers from these regional or national population database to construct sample analogue of moment conditions.

For the supply side, we assume there are $F = 5$ suppliers in national market and each produces the same number $J_f = 4$ of products. The same set of products are sold in the all regional markets. Thus, there exist $J = J_f \times F = 20$ products in national as well as each regional market.

The cost function of product j in market m is defined as

$$mc_j^m = x_j \gamma + \omega_j^m, \tag{27}$$

[5]We do not check the effect of sample sizes n and N because (1) the econometrician usually does not have control over them anyway, and (2) we assume those collecting relevant data to use large enough sample size.

where the unobserved cost shifters ω_j^m are random draws from $N(0,1)$. All suppliers have the same form of cost function. We set the cost side parameter to be $\gamma = 1.5$.

To estimate $\alpha = 1.0$, $\beta_0 = 1.0$, $\beta_1 = 0.5$ and $\gamma = 1.5$, we need instruments. We construct three instruments from $\mathbf{X}$ for the product j produced by f: x_j itself, the sum of x_k within the firm f except x_j, and the sum of x_k over the other firms than f, as BLP (1995) proposed. These three instruments are commonly used for both demand side and supply side.

As stated, we use the model-calculated conditionally true share as observed share, or we use $\mathbf{s}^0$ and not $\mathbf{s}^n$ to uncover the effect of the micro moment, without confounded by the sampling error in market share. For the same reason, we use $\boldsymbol{\eta}^0$ instead of $\boldsymbol{\eta}^N$ for the micro moment matching. Note we assume that the distribution of each regional variable is common to all the regions and moreover all drawn variables are independent.

For the additional micro moment conditions, we use two types: (1) the average of ν of consumers who choose from a set $\mathcal{Q}_1$ of the products priced higher than the national average price, (2) the average of ν of consumers who choose a set $\mathcal{Q}_2$ of the products with characteristic x_j greater than the national average. Then we construct the additional moment $\mathbf{G}_T^a(\boldsymbol{\theta}_d, \mathbf{s}^0, \mathbf{P}^R, \boldsymbol{\eta}^0)$ as in (25).

The objective function to minimize is $Q_{M.T} = \mathbf{G}_{M,T}'\mathbf{W}\mathbf{G}_{M,T}$, where

$$\mathbf{G}_{M,T} = \begin{pmatrix} \mathbf{G}_M^d(\boldsymbol{\theta}_d, \{s^0\}, \{P^R\}) \\ \mathbf{G}_M^c(\boldsymbol{\theta}, \{s^0\}, \{P^R\}) \\ \mathbf{G}_{M,T}^a(\boldsymbol{\theta}_d, \{\mathbf{s}^0\}, \{P^R\}, \boldsymbol{\eta}^0) \end{pmatrix},$$

and $\mathbf{W}$ is the weight. We use 8×8 identity matrix E as the weight, which may not give the estimator whose variance is asymptotically efficient. As with the framework of BLP (1995), we choose downhill simplex method as minimizing method and set the true value as the initial value.

We examine if and to what extent the additional moment conditions with different number T of consumer draws improve the estimation as the number of markets M increases. We use fixed $R_m = 100$ and $T = \{0, 50, 100, 500, 1000\}$ as the number of consumer draws for each market size $M = \{1, 5, 10, 20\}$. Note that $T = 0$ means that additional moments are not calculated."

Tables 1 and 2 shows the result for the averages and standard errors of the estimated parameters of $\alpha(1.0)$, $\beta_0(1.0)$, $\beta_1(0.5)$ and $\gamma(1.5)$ with/without the additional micro moment conditions for 100 Monte Carlo repetitions. Each column and each row respectively corresponds to the number of markets and the number T of consumer draws to construct those micro moments. The standard errors are in each parenthis below the corresponding means.

As expected, accuracy of the estimate measured in terms of the corresponding standard error improves as the number of market increases for both with and without additional moments. We also noticed that there are persistent biases, albeit small, in the estimates without micro moments.

For β_1, which measures consumer heterogeneity, the standard error with the micro moments is noticeably smaller than that without for the same number of markets. As expected, however, the ameliorating effects of the micro moments become weaker as the number of markets glow larger. For example, at the number of markets $M = 5$, the standard error of β_1 is greatly improved from 0.237166 with without the micro moments ($T = 0$) to 0.070259 with the micro moments constructed from the consumer draws of $T = 1000$. On the other hand, when we set the number of markets to be $M = 20$, the reduction of the standard error associated with β_1 is relatively small: from 0.109436 to only 0.060952. Although the improving effect of the micro moments diminishes, it seems to remain even when the number of markets increases.

When we calculate the micro moments with relatively few number of consumer draws at $T = 50$ or 100, the estimates are worse than that without micro moment for each number of markets. For instance, at $T = 50$, even when we set the number of markets to be $M = 20$, the average of the estimates is 0.522409 relative to the average of the estimates of 0.514595 without the micro moments. Only when we use $T = 500$ or greater consumer draws to calculate the micro moments, the standard error of the estimators starts to improve in relation to the standard errors computed without the micro moments.

For parameters other than β_1, it seems that the estimates are slightly improved with large T and regardless of the number M of markets. This is contrary to the results of Myojo and Kanazawa (2012). There the improving effects of the parameter estimates does not seem to occur in parameters other thanβ_1. Because the added moment is about consumer demographics ν, it is reasonable that additional moment conditions improves the estimate of β_1 greatly, but not for other estimates. The only reason we can come up is that the stability of the estimate of β_1 helps the estimation of other parameters. This needs to be investigated further.

From figures 1, 2, 3 and 4, the asymptotic normality starts to take hold for all the parameter estimates as the number of markets increases where the additional moment condition is available. Table 3.1 shows that both Jarque-Bera and Shapiro-Wilk tests do not reject the normality hypothesis of the estimates with $M = 20$, $R_m = 100$ and $T = 1000$.

When we can reasonably assume the asymptotic normality of the parameter estimates at $M = 20$, we can test whether the standard error of a estimator with $T = 1000$ is significantly smaller than that without additional moments. Table 4 shows the result of the test. From the result, accuracy improving effects of the additional moments remains at $M = 20$ and beyond.

Table 1: Monte Carlo Simulation Results of (α, β_0) for Multiple Markets Model With/Without Additional Moment
100 repetitions, with no sampling error ($J = 20$, $n = 10,000$, $n_m = n/M$, $R_m = 100$ each)

# of Consumer Draws (T)	α (1.0) # of Markets (M) 1	5	10	20	# of Consumer Draws (T)	β_0 (1.0) # of Markets (M) 1	5	10	20
0	1.26994	1.03313	1.01645	0.97508	0	1.52852	1.07361	1.03904	0.95806
	(1.03294)	(0.37532)	(0.204981)	(0.158212)		(1.86754)	(0.690521)	(0.369034)	(0.283556)
50	1.31111	1.06338	1.04777	1.00239	50	1.57892	1.13397	1.09765	1.01297
	(0.875849)	(0.374163)	(0.256261)	(0.214051)		(1.60783)	(0.681033)	(0.454746)	(0.371711)
100	1.31559	1.0212	1.03771	0.984661	100	1.59892	1.05349	1.07691	0.98192
	(0.867226)	(0.290003)	(0.273552)	(0.206735)		(1.59301)	(0.530612)	(0.492299)	(0.368529)
500	1.31444	1.02442	1.04496	0.991862	500	1.58161	1.0548	1.08701	0.987647
	(0.825948)	(0.243638)	(0.191984)	(0.141564)		(1.517)	(0.455524)	(0.353673)	(0.257662)
1000	1.29778	1.03497	1.0447	0.999625	1000	1.54479	1.07235	1.08683	0.999648
	(0.823522)	(0.247841)	(0.191337)	(0.128913)		(1.51631)	(0.464049)	(0.354719)	(0.237586)

Standard errors of each repetitions in the parenthesis.

Table 2: Monte Carlo Simulation Results of (β_1, γ) for Multiple Markets Model With/Without Additional Moment
100 repetitions, with no sampling error ($J = 20$, $n = 10,000$, $n_m = n/M$, $R_m = 100$ each)

# of Consumer Draws (T)	β_1 (0.5) # of Markets (M) 1	5	10	20	# of Consumer Draws (T)	γ (1.5) # of Markets (M) 1	5	10	20
0	0.56609	0.492358	0.524297	0.514595	0	1.44868	1.47612	1.48703	1.47739
	(0.579877)	(0.237166)	(0.172493)	(0.109436)		(0.239105)	(0.113698)	(0.0793864)	(0.0664014)
50	0.580772	0.522223	0.523877	0.522409	50	1.49485	1.48419	1.49174	1.47885
	(0.355857)	(0.21608)	(0.209301)	(0.212403)		(0.200528)	(0.126429)	(0.0952291)	(0.0929717)
100	0.593833	0.53132	0.530457	0.533023	100	1.49347	1.47787	1.49048	1.47432
	(0.32299)	(0.160593)	(0.162806)	(0.163413)		(0.222092)	(0.108769)	(0.0816977)	(0.0846689)
500	0.554067	0.505637	0.505657	0.504208	500	1.50725	1.48927	1.50225	1.48696
	(0.224106)	(0.093237)	(0.0901611)	(0.0807463)		(0.198426)	(0.0908213)	(0.0655352)	(0.0587172)
1000	0.565112	0.496941	0.506038	0.494323	1000	1.49857	1.4931	1.50303	1.49161
	(0.239662)	(0.0702588)	(0.0696279)	(0.0609522)		(0.20002)	(0.0898322)	(0.0625012)	(0.0542441)

Standard errors of each repetitions in the parenthesis.

Table 3: Test of normality of each estimators

	JB	p-value		W	p-value
α	3.606	0.164	α	0.980	0.135
β_0	4.697	0.0955	β_0	0.976	0.0705
β_1	1.677	0.432	β_1	0.983	0.234
γ	0.666	0.716	γ	0.993	0.879

Jarque-Bera test and Shapiro-Wilk test of each estimators when $M = 20, T = 1000$ and $R_m = 100$.

Table 4: Test of variances of the estimators

	F-value	p-value	df
α	1.5062	0.0214	(99,99)
β_0	1.4244	0.03998	(99,99)
β_1	3.2236	7.80E-09	(99,99)
γ	1.4985	0.02274	(99,99)

Alternative hypothesis: the variance of the estimator when $M = 20, T = 0$ is greater than that when $M = 20, T = 1000$

4 Conclusion and Discussion

In this paper, we implement Monte Carlo simulation to examine asymptotic properties of the estimator of random-coefficient logit models of demand for non-durable consumer goods under an equilibrium assumption in the presence of micro moments as the number of the examined regional markets increases. The national micro moments are manufactured from the joint distribution of demographic information of consumers choosing those products with certain discriminating attributes. We observe that adding an equilibrium assumption and the micro moments gives asymptotic normality with sharper asymptotic variance-covariance matrix, while correcting asymptotic bias reported in Freyberger (2015), an unexpected and rather surprising result.

Expanding the number of markets to examine seems a good idea at first, because it simply increases the number of consumers to be analyzed. Our Monte Carlo experiments as well as that of Freyberger (2015) show that this is not necessarily so, because there is no mechanism inherent in this increase in the number of markets to guarantee that the sampling is done randomly and doing so enables to achieve a good coverage of the national market. It seems adding "national" micro moments will correct possible bias inherent in this increase in the number of markets. Further research is needed to understand the mechanism under which this bias reduction is achieved.

References

[1] Berry, S., "Estimating Discrete-Choice Models of Product Differentiation," *RAND Journal of Economics,* 25 (1994), 242–262.

[2] Berry, S., J. Levinsohn, and A. Pakes, "Automobile Prices in Market Equilibrium," *Econometrica,* 63 (1995), 841–890.

[3] Berry, S., J. Levinsohn, and A. Pakes, "Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market," *Journal of Political Economy,* 112 (2004), 69–105.

[4] Berry, S., O. Linton, and A. Pakes, "Limit Theorems for Estimating the Parameters of Differentiated Product Demand Systems," *Review of Economic Studies,* 71 (2004), 613–654.

[5] *Consumer Expenditures and Income*, The US Bureau of Labor Statistics, April, 2007.

[6] Bresnahan, J., "Empirical Studies of Industries with Market Power," *Handbook of Industrial Organization, ed. by R. Schmalensee, and R. Willig,* vol. 2. (1989), North-Holland.

[7] Freyberger, J., "Asymptotic theory for differentiated products demand models with many markets," *Journal of Econometrics,* 185 (2015), 162–181.

[8] Imbens, G. and T. Lancaster, "Combining Micro and Macro Data in Microeconometric Models," *Review of Economic Studies,* 61 (1994), 655–680.

[9] Myojo, S. and Kanazawa, Y., "On Asymptotic Properties of the Parameters of Differentiated Product Demand and Supply Systems When Demographically-Categorized Purchasing Pattern Data are Available," *International Economic Review,* 53(3) (2012), 887–938.

[10] Nevo, A., "Measuring Market Power in the Ready-to-Eat Cereal Industry," *Econometrica,* 69 (2001), 307–342.

[11] Pakes, A., " Patents as Options: Some Estimates of the Value of Holding European Patent Stocks, " *Econometrica,* 54 (1986), 755-784.

[12] Petrin, A., "Quantifying the Benefits of New Products: The Case of the Minivan," *Journal of Political Economy,* 110 (2002), 705–729.

[13] Sudhir, K., "Competitive Pricing Behavior in the Auto Market: A Structural Analysis," *Marketing Science,* 20 (2001), 42–60.

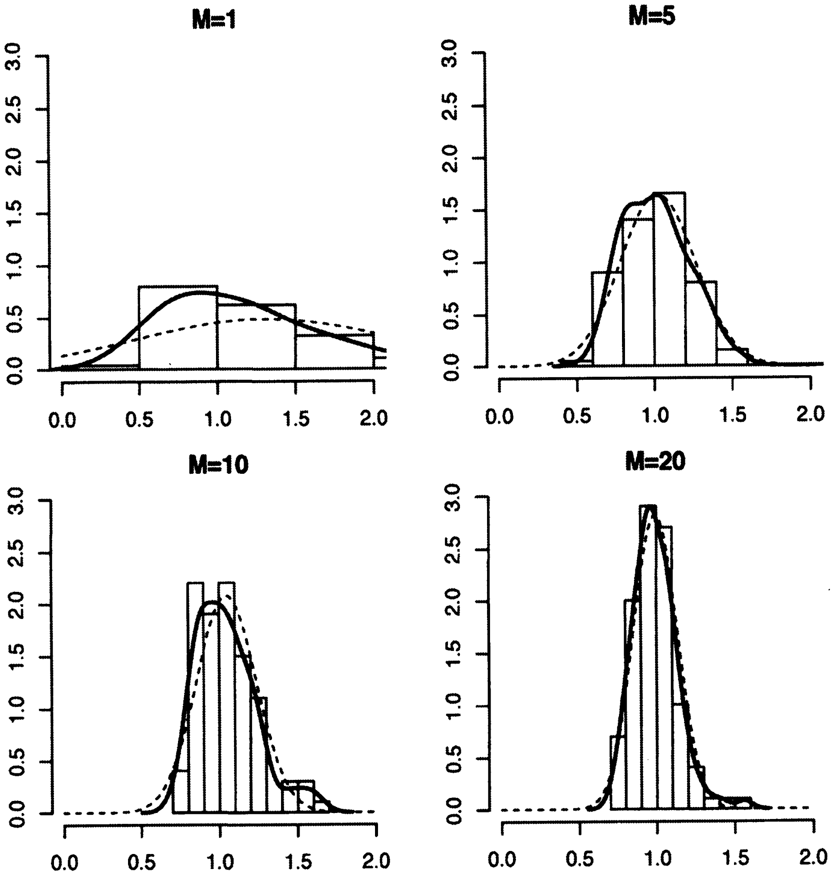

Figure 1: Histograms for the α with the additional moment (T=500) with the number $M = 1, 5, 10, 20$ of markets. Density estimates (solid) as well as the normal curves (dashed) for the α with the estimated mean and standard error are drawn in.

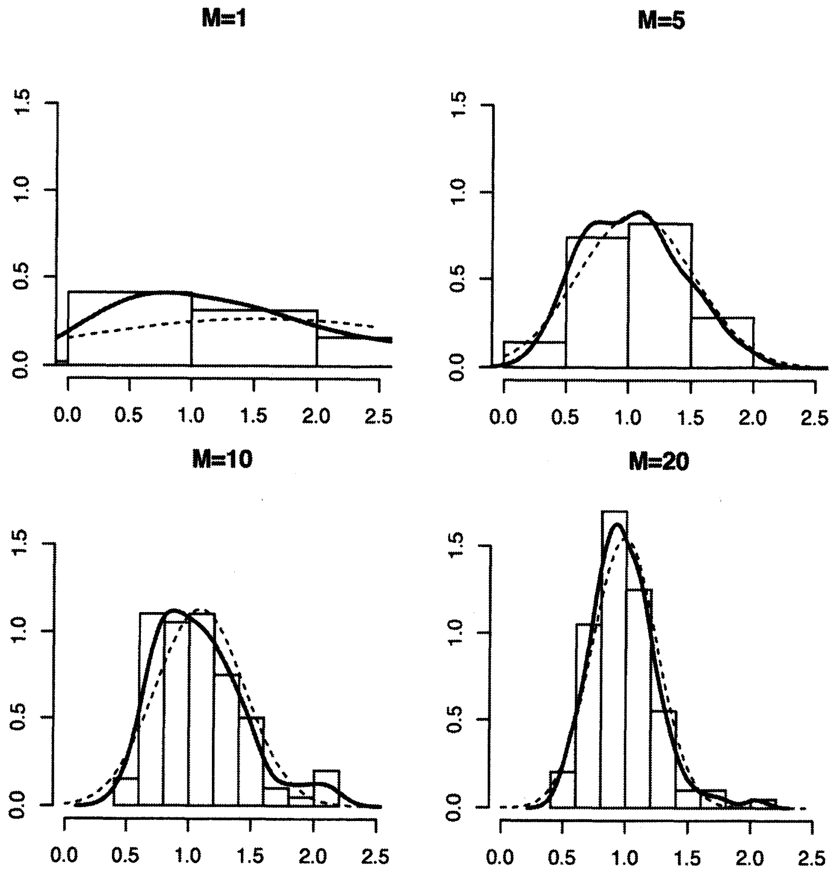

Figure 2: Histogram of β_0 with additional micro moment (T=500) with the number $M = 1, 5, 10, 20$ of markets. Density estimates (solid) as well as the normal curves (dashed) for the β_0 with the estimated mean and standard error are drawn in.

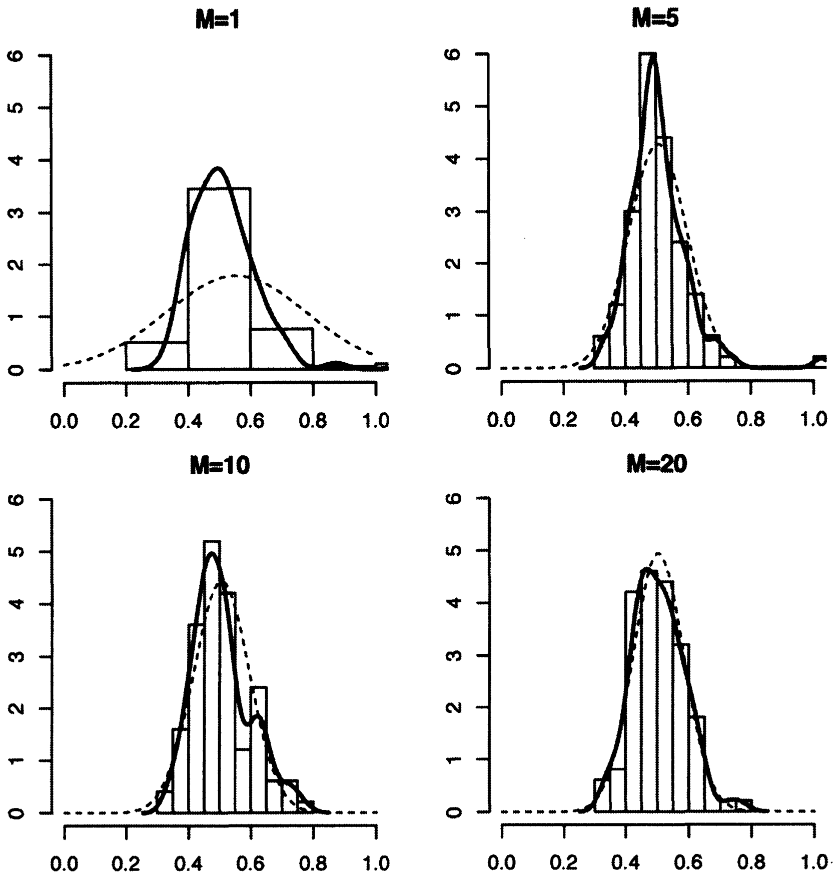

Figure 3: Histogram of β_1 with the additional moment (T=500) with the number $M = 1, 5, 10, 20$ of markets. Density estimates (solid) as well as the normal curves (dashed) for the β_1 with the estimated mean and standard error are drawn in.

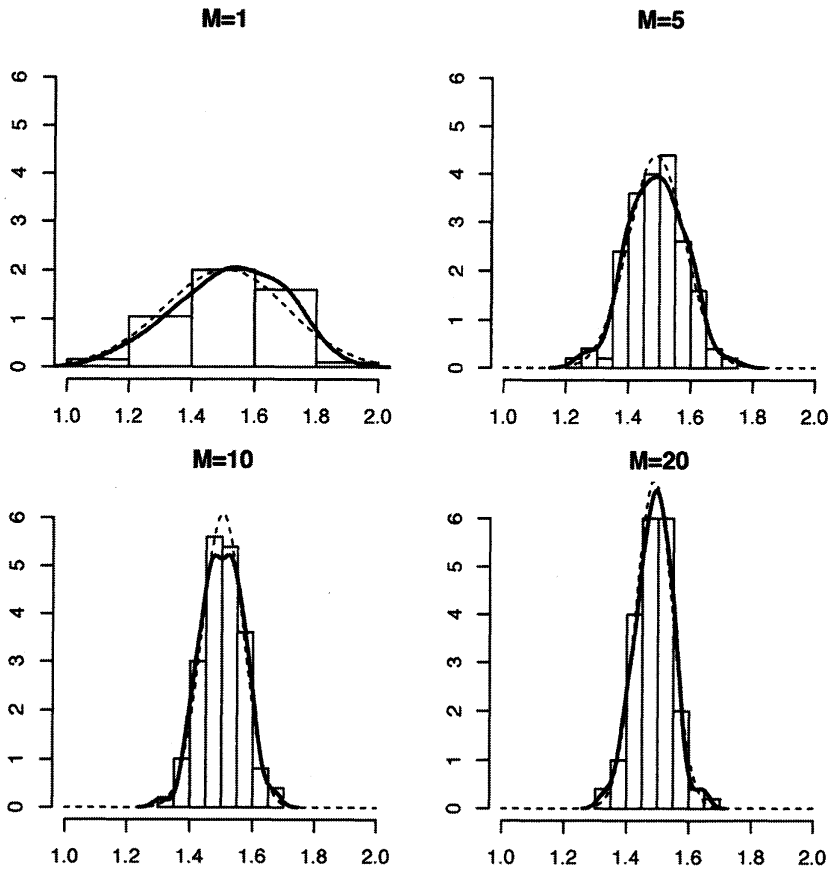

Figure 4: Histogram of γ with the additional moment (T=500) with the number $M = 1, 5, 10, 20$ of markets. Density estimates (solid) as well as the normal curves (dashed) for the γ with the estimated mean and standard error are drawn in.